Geospatial Knowledge Management

Jonathan W. Lowe

The industry has transformed huge quantities of data into information. Now the challenge is making the leap from information to knowledge — and wisdom.

Working on a geospatial project without hearing the words "integration" or "interoperability" is almost impossible these days. Another popular phrase (derived from the "data-to-wisdom pyramid") is stepping up to join them: "knowledge management."

The data-to-wisdom pyramid's broad foundation of data is topped by a layer of information, then a slice of knowledge, and finally a pointy cap of wisdom (see Figure 1). For decades the geospatial industry has been busily building our pyramid's ground floor. We've also become quite good at turning our data into information, typically by organizing it thematically in spatial databases. Our built-up data and information foundation is now supporting an emerging discipline of knowledge management that integrates, distills, and analyzes information sources to support strategic decision making.

This column investigates a few of the tools of geospatial knowledge management: semantic interoperability, rules-based data discovery, and object-based generalization, areas where small research teams and nimble companies are defining the leading edge.

Ordnance Ontology

The UK's Ordnance Survey (OS) maintains a highly detailed geospatial dataset called MasterMap that includes every building footprint in the United Kingdom (see "Ordnance Survey's MasterMap," Net Results, March 2005). Although today OS identifies each of these millions of polygons simply as "Building," market factors may lead the agency to expand its model from storing only physical form to including both form and function. For example, in a future release OS may describe its "Building" polygons with attributes such as "Community Center," "Shopping Center," "Stadium," or "Church."

To add such functional attribution to its dataset, OS has some challenges to surmount, such as cases where a building serves different functional purposes on different days (e.g., a sports center that's a church on weekends). And a small group of OS researchers — Catherine Dolbear, John Goodwin, and Hayley Mizen — also anticipate a problem that could prevent uptake of Ordnance Survey data even after the challenges of capturing a building's core functional attributes are overcome. They recognize that, because there is no way of predicting all the possible uses of OS data, MasterMap's conceptual model won't match all of their potential users' conceptual models. Having to adapt between incompatible models will impede users' ability to exploit OS data to its full potential.

For instance, what if an emergency responder needs to identify buildings that could serve as temporary shelters after a flood? To a responder, the "Emergency Shelter" attribute includes criteria such as:

- Provides shelter to people;
- Has capacity for 500 people;
- Is located near main roads and transport links;
- Is close to a large supermarket;
- Is located at least one-half mile from the disaster event (e.g., the flood zone);
- Is readily accessible to public and local authorities; and
- Has facilities such as telephones, restrooms, seating, and heating.

The OS dataset models real-world objects that satisfy these criteria, but when the responder directly searches for "Emergency Shelter," she probably won't find any buildings identified by that specific attribute. Her search will fail because OS functional attribute names can't anticipate temporary usage of buildings (such as stadiums or churches) as emergency shelters, even if they are perfect candidates for that use during a particular disaster. Similarly, although a building may typically function as a shopping mall, it

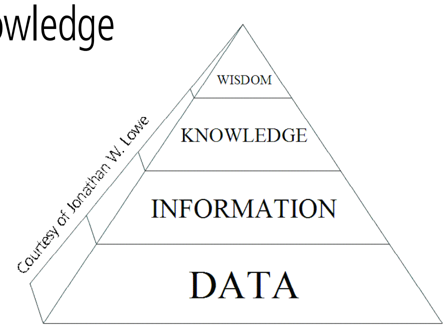
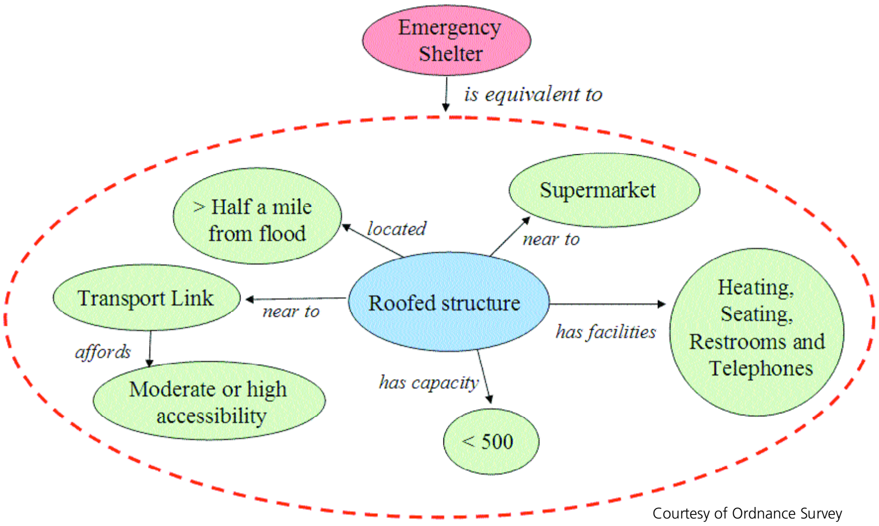


Figure 1. The data-to-wisdom pyramid illustrates the challenge of modern life — refining large volumes of raw data to acquire the more valuable information, knowledge, and (most rarely) wisdom that rest upon that super-abundant but heterogeneous foundation.

Net Results columnist **Jonathan W. Lowe** covers the role of emerging technologies in the exchange of spatial information. Lowe is on assignment in London to assist the United Kingdom's Department for Environment, Food, and Rural Affairs with its geospatial repository. He can be contacted at info@giswebsite.com.



Courtesy of Ordnance Survey

Figure 2. Emergency shelter ontology: Ordnance Survey's semantic solution to interoperation of differing conceptual frameworks and a common spatial data source is to transform the domain expert's conceptual model into a logical ontology. Crown copyright 2005.

might also fit the definition of an emergency shelter, a terrorist target, or a commercial real estate listing, depending on the user. How will OS model its data to serve a disparate group of stakeholders, including emergency responders, public safety groups, real estate agents, and others?

Crafting a Common Language

Recognizing that the emergency responder's search would fail within today's knowledge management context, the OS research team is developing an alternative. The team's new method uses a knowledge representation approach that describes the semantic meaning of the geospatial datasets via ontologies (see "A Geospatial Semantic Web," Net Results, June 2005).

The first step in the team's approach is to capture the responder's definition of an emergency shelter in her own words. An ontology specialist then turns the responder's list of criteria into an emergency shelter ontology (see Figure 2). The ontology is a formal structure naming concepts and the relationships between them, and has similarities to methodologies such as Unified Modeling Language, but is designed to capture deeper semantics and is capable of being tested for logical consistency.

The ontology then provides a bridge between the language and meaning of an emergency responder and the more generic structure of the OS spatial dataset. For this to take place, the OS data must have its own ontology providing a reference to its data's common terms and relationships (and the agency is actively acquiring requirements for an ontology covering what it calls the "topographic domain"). For the building polygons' functional uses, for instance, the ontology might capture that "stadiums have capacity for at least 300 people" or "civic buildings have restroom and telephone facilities" or "town halls are roofed structures."

Given these two formalized conceptual frameworks — the responder's and OS' ontologies — the basis for interoperation becomes possible. The responder needs buildings with facilities such as restrooms and telephones; by inference, the OS ontology would identify which buildings in its list of functional building uses have such facilities. The responder needs buildings with roofs; by inference, the OS ontology would identify those as well. Because both conceptual frameworks are in formal structures — or ontological "languages" — the interoperability enabling a query for emergency shelters could be handled automatically by an inference engine (see Figure 3). Although there won't necessarily be a one-to-one match, OS has successfully tested this approach to answer the question, "Where are all the buildings that could be used as emergency shelters during the Northam flood event?" (see Figure 4). The resulting list gives our emergency responder a manageable starting point

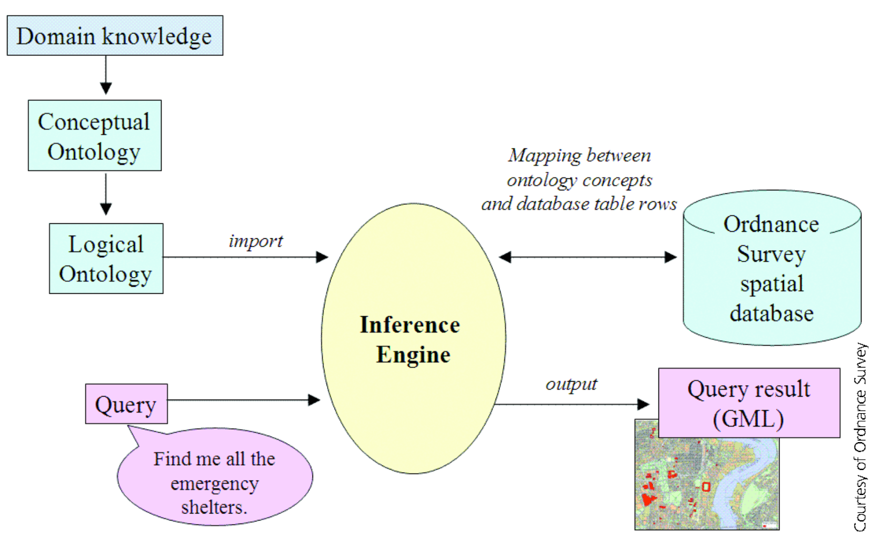


Figure 3. Semantic architecture: With a logical ontology, Ordnance Survey can join concepts to instantiated data using an inference engine to field questions about emergency shelters, even though the term doesn't exist explicitly do so in the Ordnance Survey dataset.

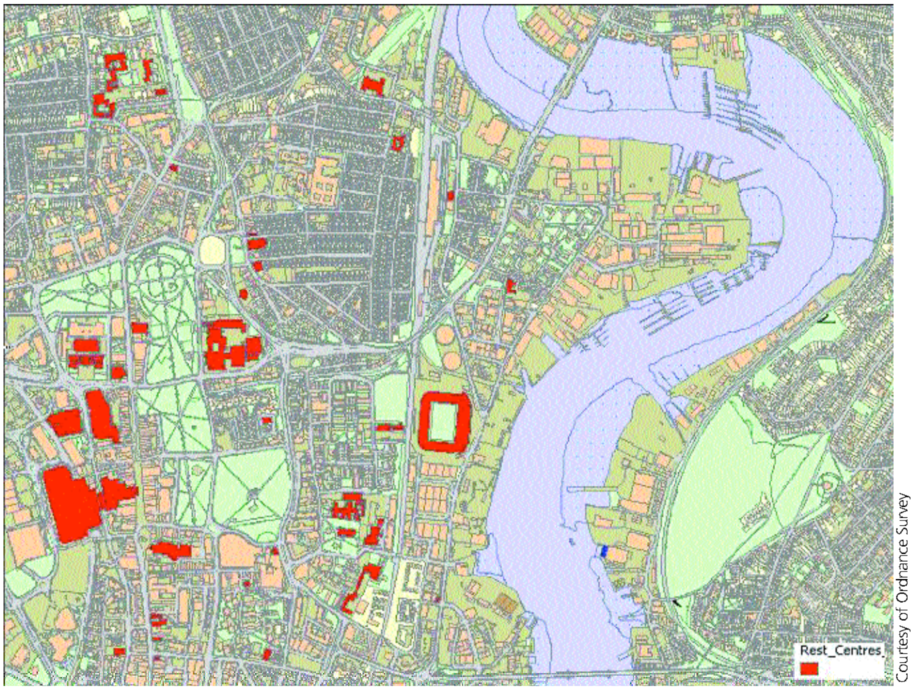


Figure 4. A semantically supported Ordnance Survey dataset can answer the query, "Find all the buildings that could be used as emergency shelters during the Northam flood event," even if none of the buildings are specifically attributed as "Emergency Shelters."

for deciding which buildings she will put into use as emergency shelters.

Rules and Refinement

OS is not alone in harnessing knowledge management concepts for today's geospatial challenges. Geospatial solution provider Laser-Scan was among the exhibitors at OS' recent "Terra future" conference, where Chief Scientist Paul Watson and Product Manager Chris Tagg demonstrated their Radius Studio product. (Radius Studio is currently undergoing beta testing but will be generally available in January 2006.) Whereas OS uses ontologies to make searches more flexible, Laser-Scan's Radius Studio uses ontologies to support rules-based data integrity and fix-up procedures.

Radius Studio exemplifies our industry's climb up the data-to-wisdom pyramid in that it enables comparison of standalone datasets against one another. Most practitioners are used to checking an individual dataset's quality — checking its geometry for unclosed polygons, overshoots, kickbacks, and a host of other errors. Checks can also include geometry comparison and topologic integrity checks within the same thematic layer, such as confirming that contiguous boundaries never overlap. (GeoServer, a free, open-source product, provides such rules-based layer-specific integrity checks, for instance.) As important to data quality as they are, these quality tests operate at the information level of the pyramid. Laser-Scan is helping users step up another level to manage knowledge by providing comparisons not just within layers, but between them.

For instance, Radius Studio can store business rules such as "Buildings must be completely contained by zoning boundaries," and then confirm the rules with commands such as "Check each Building object to confirm that at least one Zoning object's geometry contains that Building object's geometry." Radius Studio provides non-technical users with an organized interface (see Figure 5) for building a set of rules that Laser-Scan hopes will empower domain experts (for instance, hydrologists or city planners) to work directly with spatial data rather than through technicians who may incorrectly interpret the experts' domain rules. Such domain experts may never realize that they are building ontologies, but that's exactly what is happening behind the scenes.

Not only does Radius Studio provide a user-friendly window into ontology design, but it can perform "data-discovery" queries against a collection of layers, even without prior knowledge of inter-layer relationships. In discovery mode, Radius Studio looks for repeated

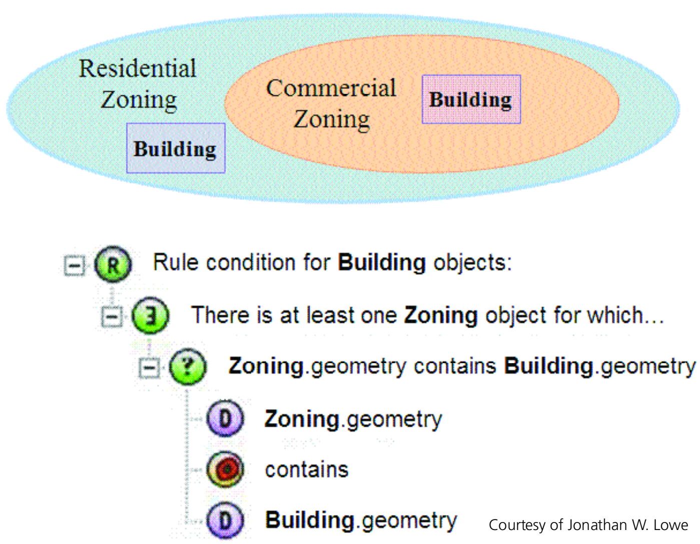


Figure 5. Laser-Scan's Radius Studio can check and fix existing spatial datasets or validate incoming updates to clean datasets based on contextual rules. Radius Studio lets nontechnical users express the spatial relationships in this buildings and zoning diagram using nested logical statements formatted for easy interpretation.

patterns such as "The majority of points from the 'Fire Stations' layer are contained by polygons from the 'Building Footprints' layer," or "Ninety-five percent of the lines from the 'Streets' layer have nodes matching the edges of a containing polygon from the 'Pavement' layer." Then it's up to the data administrator to verify whether these relationships represent intended rules or are simply coincidental.

To perform respectably, Radius Studio relies on Laser- Scan's topology engine (also available in an associated Laser-Scan product, Radius Topology), which converts geometric objects into topologic structures for speedy processing when testing complicated relationship rules. The Laser-Scan product line is designed for use with Oracle spatial database implementations.

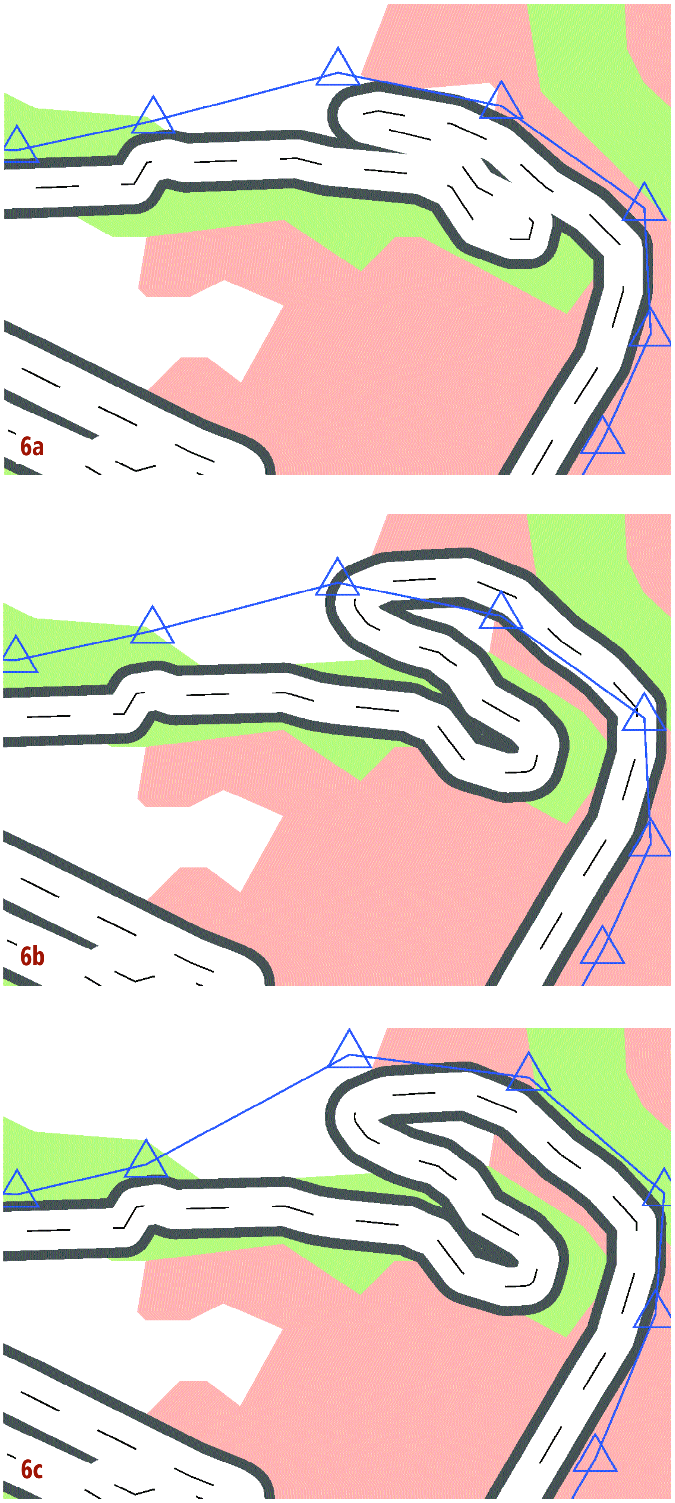
The Big Picture

Knowledge management isn't limited to the interoperability of diverse applications and related datasets. It also involves distilling the essence from a large collection of information, as a data warehouse is designed to summarize and track trends in a companion transactional database. Spatially, generalization means reducing the detail of a layer or entire data collection without losing its original overall meaning. Spatial generalization technology is also climbing the data-to-wisdom pyramid, as evidenced by another Laser-Scan product called Clarity.

At the base of the data-to-wisdom pyramid, the age-old generalization problem is to remove a maximum number of vertices from a line or polygon without radically altering its general shape. Today, such algorithms are better characterized as tools for simplification or "weeding" rather than generalization, and remain useful as methods for improving rendering speed or reducing the size of a single layer for use on a handheld device with limited storage.

At the information level, generalization should thin out data without changing its layer-specific topologic integrity. Generalizing a street network layer should not break any of the node-to-node connections between originally joined-up street segments, for instance. But solutions to problems at the data and information levels are fairly well-established.

At the knowledge level, however, generalization can be dangerous if the algorithm is not aware of relationships between layers. For instance, without reference to relationship rules between



Figures 6a–6c. The original data **(a)** show the road below the line of poles. Generalization alters the road for clear recognition at larger scales, but changes its relationship to the poles **(b)**. Laser-Scan's generalization solution preserves the original relationship between the road and pole objects **(c)**.

Images courtesy of Jonathan W. Lowe

Figures 7a–7c. A Laser-Scan "mesoagent" generalizes the original cluster of buildings **(a)** to remove extraneous detail **(b)**, then combines sub-groups of the cluster for clarity of display at a smaller scale **(c)**.

roads and utility lines, a line of telephone poles that curves around the perimeter of a road polygon could be generalized such that it intersects the road or appears on its opposite side (**see Figure 6**).

Another generalization challenge involves combining similar features, actually reducing the number of records in a layer. For instance, at 1:2,500 scale a map can comfortably include each footprint in a cluster of 50 small buildings, but at 1:25,000 scale the same cluster could be represented by just 10 exaggerated and tightly clustered buildings to suggest the same sort of settlement without cartographic confusion (**see Figure 7**).

Laser-Scan's approach to these sophisticated generalization challenges relies on a relationship rule base, similar to that used by Radius Studio, and algorithms (that it calls "agents") to generalize collections of data at both "micro" and "meso" levels. Microagents simplify geometries independently of other objects, while mesoagents evaluate groups of data in the same region, applying user-established rules and priorities to achieve acceptable cartographic results.

Laser-Scan's customers report that the results of generalization at the knowledge management level still require review by human eyes for specific cases, but that the overall time savings can change the entire map-production cycle. For instance, the French national mapping agency — the Institut Geographique National (IGN) — produces a variety of map products at different scales, originally requiring about 1,200 hours per map sheet to derive a 1:50,000-scale map series from their more granular base data using manual generalization techniques. Using Laser-Scan's Clarity solution, IGN reduced the same process to 50 hours of automated processing and 100 hours of manual reworking per map sheet.

Word to the Wise

OS and Laser-Scan provide examples of an emerging industry shift up the pyramid from data to information to knowledge management. The trend is enabled by the superabundance of data, the organized storage of that data in relational databases, and the growing evidence that integrating corporate information sources, both human and electronic, results in a competitive advantage over less interoperable players. Is it premature to wonder what the leap from knowledge to wisdom will entail? May we all still be around to see it. ⊕

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION

(Required by 39 USC 3685)

1. Publication Title: Geospatial Solutions
2. Publication Number: 1529-7403
3. Filing Date: 9/23/05
4. Issue Frequency: Monthly
5. Number of Issues Published Annually: 12
6. Annual Subscription Price: $67.00
7. Complete Mailing Address of Known Office of Publication: 306 West Michigan Street, Suite 200, Duluth, St. Louis County, Minnesota 55802-1610 Contact Person: James Tuttle Telephone: 218-279-8860
8. Complete Mailing Address of Headquarters or General Business Office of Publisher: 275 Grove Street, Suite 2-130, Newton, MA 02466
9. Full Names and Complete Mailing Addresses of Publisher: Dana Fisher, 201 Sandpointe Avenue, Suite 600, Santa Ana, CA 92707 **Editor:** Scottie Barnes, 86718 Territorial Rd., Veneta, OR 97487 **Managing Editor:** Amy Stankiewicz, 7500 Old Oak Blvd., Cleveland, OH 44130
10. This publication is owned by: Questex Media Group, Inc., 275 Grove Street, Suite 2-130, Newton, MA 02466. The sole shareholder of Questex Media Group, Inc. is: QMG Holdings, Inc., 275 Grove Street, Suite 2-130, Newton, MA 02466
11. Questex Media Group, Inc. is the Mortgagor under a Credit Agreement dated May 23, 2005, with various lenders as named therein from time to time. The agent for the lenders is: Bank of Montreal, Chicago Branch, as Administrative Agent, 115 South LaSalle Street, Chicago, IL 60603. Holders of 1.0% or more of Questex Media Group, Inc. Mortgages or Other Securities as of July, 2005 are as follows: Audax Private Equity Fund, L.P. 101 Huntington Avenue Boston, MA 02199 Audax Private Equity Fund II, L.P. 101 Huntington Avenue Boston, MA 02199 Audax Co-Invest, L.P. 101 Huntington Avenue Boston, MA 02199 Audax Trust Co-Invest, L.P. 101 Huntington Avenue Boston, MA 02199 Audax Special Purpose Co-Invest, L.P. 101 Huntington Avenue Boston, MA 02199 AFF Co-Invest, L.P. 101 Huntington Avenue Boston, MA 02199 Bank of Montreal 3 Times Square, 29th Floor New York, NY 10036 Ares Enhanced Loan Investment Strategy, Ltd Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Ares Enhanced Loan Investment Strategy II, Ltd Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Ares Total Value Fund, LP Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Ares IV CLO Ltd Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Ares V CLO Ltd Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Ares VII CLO Ltd Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Ares VIII CLO Ltd Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Ares IX CLO Ltd Ares Management LLC 1999 Avenue of the Stars #1900 Los Angeles, CA 90067 Friedberg Milstein Private Capital Fund I C/O Friedberg Milstein LLC 6 East 43rd Street New York, NY 10017 FM Leveraged Capital Fund I 6 East 43rd St New York, NY 10017 Green Lane CLO Ltd C/O Guggenheim Investment Mgmt LLC 135 East 57th Street, 9th Floor New York, NY 10022 GSC Partners CDO Fund III Limited C/O GSCP (NJ) LP as collateral Manager 300 Campus Drive, Suite 110 Florham Park, NJ 07932 GSC Partners CDO Fund V Limited C/O GSCP (NJ) LP as collateral Manager 300 Campus Drive, Suite 110 Florham Park, NJ 07932 GSC Partners Gemini Fund Ltd C/O GSCP (NJ) LP as collateral Manager 300 Campus Drive, Suite 110 Florham Park, NJ 07932 LFC2 Loan Funding LLC C/O LaSalle Bank NA 135 South LaSalle Street, Suite 1625 Chicago, IL 60603 Midland National Life Insurance Co C/O Guggenheim Investment Mgmt LLC 135 East 57th Street, 9th Floor New York, NY 10022 North American Co for Life and Health Ins C/O Guggenheim Investment Mgmt LLC 135 East 57th Street , 9th Floor New York, NY 10022 Wells Fargo Foothill Inc 2450 Colorado Avenue, Suite 3000W Santa Monica, CA 90404
12. Does Not Apply
13. Publication Title: Geospatial Solutions
14. Issue Date for Circulation Data Below: August 2005

15. Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
A. Total Number of Copies	**28,600**	**28,388**
B. Paid and/or Requested Circulation		
1. Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541	25,688	25,557
2. Paid In-County Subscriptions Stated on Form 3541		
3. Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution (Includes non-USPS requester distribution)	1,520	1,671
4. Other Classes Mailed Through the USPS		
C. Total Paid and/or Requested Circulation	27,208	27,228
D. Free Distribution by Mail		
1. Outside-County as Stated on Form 3541	522	495
2. In-County as Stated on Form 3541	0	0
3. Other Classes Mailed Through the USPS	0	0
E. Free Distribution Outside the Mail	771	569
F. Total Free Distribution	1,293	1,064
G. Total Distribution	28,501	28,292
H. Copies Not Distributed	99	96
I. Total	28,600	28,388
J. Percent Paid and/or Requested Circulation	95.5%	96.2%

16. Publication required. Will be printed in the NOVEMBER issue of this publication
17. Name and Title of Editor, Publishers, Business Manager, or Owner:
Kristin J. Meyer, Circulation Director **Date:** 9/23/05

I certify that the statements made by me above are correct and complete.